Modeling the Decline in English Passivization

Liwen Hou
College of Computer and
Information Science
Northeastern University
lhou@ccs.neu.edu

David A. Smith
College of Computer and
Information Science
Northeastern University
dasmith@ccs.neu.edu

Abstract

Evidence from the Hansard corpus shows that the passive voice in British English has declined in relative frequency over the last two centuries. We investigate which factors are predictive of whether transitive verb phrases are passivized. We show the increasing importance of the person-hierarchy effects observed by Bresnan et al. (2001), with increasing strength of the constraint against passivizing clauses with local agents, as well as the rising prevalence of such agents. Moreover, our ablation experiments on the Wall Street Journal and Hansard corpora provide support for the unmarked information structure of 'given' before 'new' noted by Halliday (1967).

1 Introduction

From the Hansard corpus of British parliamentary proceedings (Alexander and Davies, 2015), we observe that the passive voice has declined in usage frequency over the last two centuries. The early 19th century saw more frequent usage of this voice compared to the late 20th century: As shown in Figure 1, while the passive was used in approximately 8% of two-argument clauses in the 1830s for example, it was used in less than 6% of such clauses in the 1990s. Following Bresnan et al. (2001), we exclude short passives that contain no "by" phrase to focus on those two-argument clauses where active and passive voices are in direct competition.

Four corpora (LOB, F-LOB, Brown, and Frown) are used by Mair and Leech (2006) to argue that the passive decreased in frequency in written English between the 1960s and the 1990s to match the norms of spoken English. While our analysis does not rule out this effect of converging registers, we provide evidence for additional factors in the evolution of the English passive.

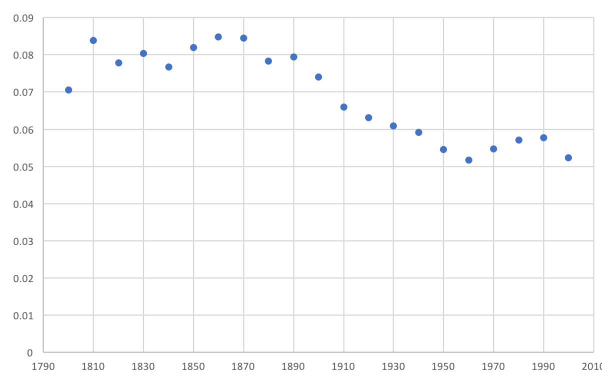

Figure 1: Proportions of passivized two-argument finite verb phrases in the British Hansard over 200 years

In this paper, we investigate the causes of the passive's decline as follows: First, we identify features that are predictive of whether a verb phrase is passivized by building a logistic regression model using features suggested in the literature. After identifying important explanatory variables for predicting passivization, we use the British Hansard corpus to investigate the change in average value of each feature over time to find explanations for the decline in passivization and then discuss the changes undergone by feature weights over time. We show the rising importance of person-hierarchy effects in English passivization noted by Bresnan et al. (2001), with increasing strength of the constraint against passivizing clauses with local (i.e. first- or second-person) agents and the increasing prevalence of such agents.

The majority of work on diachronic syntax has relied on manual annotations, and computational techniques in historical linguistics have mostly focused on phonology, morphology, and the lexicon (Lieberman et al., 2007; Ellison and Kirby, 2006; Bouchard-Côté et al., 2013, for example). One additional goal of this paper, therefore, is to employ automated methods to analyze the factors that affect passivization and that explain its decreasing frequency in English over the last two centuries.

2 Data

2.1 The British Hansard

To identify diachronic trends, we use the Hansard corpus (Alexander and Davies, 2015), which is a digitized version of two centuries of debates that took place in the British Parliament starting from 1803. We divide the data according to the decade in which each speech was given. When fitting models, we discard the decades prior to 1830 due to the small amount of data from those years.

The number of two-argument constructions and the number of words in each decade from 1830 to 1999 is shown in Table 1.

Decade	Words	Transitive Verbs
1830s	16,427,918	404,685
1840s	15,464,589	403,245
1850s	16,838,010	392,244
1860s	16,850,076	428,572
1870s	19,922,209	460,881
1880s	30,082,916	698,427
1890s	22,489,078	546,254
1900s	24,835,231	719,629
1910s	29,375,435	1,028,534
1920s	20,501,261	818,339
1930s	35,428,497	1,598,164
1940s	32,802,372	1,565,450
1950s	31,907,582	2,091,740
1960s	36,915,775	2,668,257
1970s	37,551,800	3,015,740
1980s	40,065,521	3,516,300
1990s	33,978,717	3,396,495

Table 1: The number of words and the number of two-argument actives/passives per decade

After parsing the text of the parliamentary debates using version 3.6.0 of the Stanford dependency parser (Manning et al., 2014; De Marneffe et al., 2006), we detect passive verb phrases by screening for two dependency relation types: those labeled with "auxpass" and with "nsubjpass". (In transitive constructions, passives can also be detected by screening relations that have the label "nmod:agent".) To identify the agent in a passive construction, we focus on the labels ending with "agent". To identify the subject of each verb, we use the labels "nsubjpass" and "nsubj". Finally, to identify the object in an active construction, we make use of the "dobj" relation. Although not all demoted subjects in passive constructions are agents, and not all promoted objects are patients, we use the terms "agent" and "patient" to refer to the former and the latter respectively. The aforementioned identification process yields approximately 26 million two-argument verb phrases in total, of which roughly 1.5 million are passives.

2.2 Evaluating Parser Accuracy

We have verified 300 clauses to ensure that the Stanford parser is sufficiently reliable for processing language from the Hansard corpus despite the time period covered by this corpus. The results of our manual verification process are listed in Table 2.

verb valency & voice	argument acc.	accuracy
2-argument actives	84%	95%
2-argument passives	88%	94%
active intransitives	86%	90%
short passives	92%	94%

Table 2: Manual evaluation of parser accuracy: In the middle column, a parse was considered accurate only when all arguments and the voice were correctly identified. In the last column, only valency and voice had to be correct.

We randomly sampled 100 verb phrases that were identified by the parser as being two-argument actives; of those, it correctly identified the voice and valency in 95 cases, but in only 84 cases were both arguments correctly identified as well. We also randomly sampled 100 verb phrases from the pool identified by the parser as two-argument passives; of those, it correctly identified the voice and valency in 94 cases (and both of the arguments in 88 samples). Finally, we sampled 100 verbs (50 actives and

50 passives) identified by the parser as having only one argument; it was correct 92% of the time on this sample for voice plus valency (and 89% of the time for identifying the argument).

In clauses with two arguments, the most common type of parser error was incorrectly identifying some argument of the verb. In addition, there were cases where the parser decided that a transitive verb phrase had only one argument and vice versa (e.g. treating copular constructions as transitive; classifying instrumental prepositional phrases as agentive); such valency errors were more common than voice identification errors. From our sample, we have observed that the parser rarely decides on an incorrect voice, which means that passives are correctly identified as such by the parser in the vast majority of cases.

2.3 Wall Street Journal

In addition, we report our model's performance on the Wall Street Journal corpus from the Penn Treebank Project (Marcus et al., 1993; Marcus et al., 1994) and use the latter to test the significance of explanatory variables. Even though accurate constituency trees are provided in the Wall Street Journal corpus, we parsed the text with a dependency parser and processed it in a manner similar to the aforementioned procedure for the Hansard so that results from the two corpora would be comparable.

3 Modeling Passivization

We fit a logistic regression model to predict the passivization of two-argument verbs. Similar to earlier models inspired by Harmonic Grammar (Legendre et al., 1990), we start with only the constraints on the locality of agent and patient.

Obtaining better predictions (on our task of interest of predicting whether a given verb phrase is passive or active) would help us identify the most important explanatory variables for why a speaker might choose to use the passive over the active voice.

3.1 Features

The features in our model were inspired by several previous studies of English passivization.

Person Features In our simplest model, inspired by the work of Bresnan et al. (2001) on person-hierarchy effects (see §5.1), each data point consists of two binary features. The first indicates whether or not the agent is a local (i.e. first or second) person, and the other corresponds to whether or not the patient is a local person.

Pronoun Features Because the existence of person-hierarchy effects would be confounded by the fact that local persons happen to be pronouns and thus more likely to be "given" information (see below), we add two features to denote whether the agent is a pronoun and whether the patient is a pronoun. In addition to personal pronouns, we include demonstrative pronouns (i.e. "this", "that" when the part-of-speech tag is "DT") as pronouns.

Length Features The length of a constituent was reported by Wolk et al. (2013) to have an effect on predicting the dative alternation. Specifically, a double object dative has a greater likelihood of being realized in British English as well as American English (and especially the latter) when the length of the patient is longer. We therefore also consider the length of the agent and that of the patient when predicting passivization.

Taking square roots of the lengths led to better performance on development data compared to using the raw lengths, so our two length features consist of the square root of the agent's length and that of the patient's length.

Given or New Information As a proxy for given information, we add a feature corresponding to whether the agent begins with the lemma "the", "this", "that" or a pronoun, as well as another feature indicating whether the patient begins with one such word.

Relative Clauses and Wh-words We add two features indicating whether the current verb is part of a relative clause and whether it is part of a clause beginning with a wh-word.

Preceding Passives Parallel structure among successive sentences was reported by Weiner and Labov (1983) to have a significant effect on whether a sentence contains a passivized verb. In the same vein, we add two features representing preceding passives: the first indicates whether or not any of the previous five verbs was passivized, and the second indicates whether there was a passive in any of the previous five sentences.

Lemma Features Finally, we add 1,000 features representing the 1,000 most common verb lemmas, with one additional feature to catch the remaining less common verbs. In order to see the effects of having different agent and patient head words, we also add 2,002 binary features corresponding to the 1,000 most common agents and the 1,000 most common patients (along with two features to catch the remainder) from across all years.

3.2 Performance

Since, as noted above, a little over 5% of two-argument transitive clauses in the Hansard as a whole are passive, a classifier that always predicts 'active' can achieve quite high token-level accuracy. For each test set, we report the proportion of active clauses as the "baseline" accuracy. We also, therefore, report not just the raw classifier accuracy on test data but also the precision, recall, and F1 for correctly detecting passive clauses. All evaluations are the result of five-fold cross validation.

Hansard Table 3 shows the full model's performance on different decades of the Hansard corpus, with each decade treated independently.

Decade	Acc.	F1	Prec.	Recall	Baseline
1830	0.962	0.745	0.772	0.721	0.923
1850	0.959	0.736	0.767	0.707	0.920
1870	0.957	0.734	0.765	0.705	0.917
1890	0.961	0.744	0.771	0.719	0.921
1910	0.969	0.754	0.782	0.729	0.935
1930	0.970	0.740	0.775	0.708	0.939
1950	0.970	0.712	0.759	0.671	0.945
1970	0.970	0.702	0.761	0.652	0.945
1990	0.966	0.673	0.755	0.608	0.942

Table 3: the full model's performance on every other decade of the British Hansard corpus since 1830

Sorting the features in each decade by the magnitude of their coefficients, the top four features are the same in every decade of the Hansard from the 1830s to the 2000s: whether the agent is given (instead of new), the agent lemma "who", the agent lemma "which", and the verb lemma "have". (The coefficients in these four cases are all negative, which means that these features suppress passivization.)

Wall Street Journal Table 4 shows the performance of the full model on the WSJ corpus.

Accuracy	F1	Prec.	Recall	Baseline
0.964	0.395	0.733	0.270	0.957

Table 4: Performance on the Wall Street Journal

The top ten features sorted by the magnitude of their coefficients are shown in Table 5. Positive weights correspond to passivization being more likely; for example, our results show that the verb "have" is rarely passivized, while the verb "offset" is passivized relatively often.

Top Feature	Weight
Given Agent	-5.039
Lemma "which" (Agent)	-3.257
Lemma "Tenders" (Patient)	3.130
Lemma "who" (Agent)	-3.126
Lemma "have" (Verb)	-2.851
Lemma "offset" (Verb)	2.815
Lemma "rate" (Verb)	2.334
Lemma "who" (Patient)	2.168
Lemma "affect" (Verb)	2.100
Lemma "cover" (Verb)	2.027

Table 5: Top Features for the WSJ Corpus

3.3 Effects from Feature Classes

We measure the effects on performance of removing classes of features from the full model and making predictions with five-fold cross-validation. Table 6 shows the predictive accuracy and F1 score achieved on the Wall Street Journal corpus by each ablated model.

Removed Features	Accuracy	F1	p-value
Pronouns	0.964	0.393	0.196
Lengths	0.961	0.283	< 0.005
Given	0.962	0.299	< 0.005
Rel.&Wh-Clauses	0.964	0.395	0.391
Preceding Passives	0.964	0.382	0.006
Persons	0.964	0.394	0.243
Lemma Features	0.958	0.122	< 0.005

Table 6: Effects on the WSJ of removing one group of features at a time

Compared to the performance reported in Table

4, the lemmas were the feature category whose removal caused the biggest impact on performance.

We have also found that the lemma features are extremely important for predicting passivization on the Hansard corpus. For example, if we use only the top 10 instead of the top 1,000 lemma features of each type, the F1 score for the full model drops from 0.745 to 0.514 for the 1830s and similarly drops from 0.673 to 0.391 for the 1990s.

To see the effects of the other features on the F1 score more clearly, we trained the model using only the top 10 lemmas of each type and then removed each of the other feature categories in turn to measure the decrease in performance. We did this separately for different decades of the Hansard; the results obtained for the 1830s and 1990s are shown in Table 7 and Table 8.

Removed Features	Accuracy	F1	p-value
(None)	0.936	0.514	
Persons	0.936	0.512	0.380
Pronouns	0.936	0.504	< 0.005
Lengths	0.934	0.461	< 0.005
Given	0.926	0.183	< 0.005
Rel.&Wh-Clauses	0.936	0.512	0.445
Preceding Passives	0.936	0.504	< 0.005
Lemma Features	0.927	0.313	< 0.005

Table 7: Effects on the 1830s of removing individual features (with the top 10 lemmas of each type in the full model)

Removed Features	Accuracy	F1	p-value
(None)	0.949	0.391	
Persons	0.949	0.391	0.245
Pronouns	0.949	0.386	< 0.005
Lengths	0.948	0.368	< 0.005
Given	0.946	0.218	< 0.005
Rel.&Wh-Clauses	0.949	0.372	< 0.005
Preceding Passives	0.949	0.373	< 0.005
Lemma Features	0.944	0.231	< 0.005

Table 8: Effects on the 1990s of removing individual features (with the top 10 lemmas of each type in the full model)

3.4 Statistical Significance

To test the statistical significance of the contribution of individual features, we compare the full model to each smaller model from Section 3.3 using a permutation test.

To test whether one model outperforms another in a statistically significant way, we swap or keep each pair of outputs with equal probability and, in this way, generate two new series of outputs. We then measure the difference in F1 scores between these two series of predictions and repeat this procedure 200 times to generate 200 such differences. Next, we compare the true difference in F1 scores of the original models to the 200 randomly generated differences. The reported p-values are the proportions of the randomly generated differences that are as large as or larger than the true difference.

For the Wall Street Journal, we see from Table 6 that the features whose removal caused a statistically significant decrease in the F1 score were the features representing the lengths of the agent and patient, the features indicating whether the agent and patient were given (instead of new) information, the features indicating whether the preceding five verb phrases and the preceding five sentences contained passives, and finally the lemma features.

For the Hansard corpus, after using only the top 10 lemmas of each type, we see from Table 7 and Table 8 that the removal of any feature category other than the person features (and relative/wh-clause features in the 1830s) causes a statistically significant decrease in the F1 score. In particular, removing the given/new information features causes the F1 to suffer the biggest drop.

4 Changes in Feature Values

We have thus far identified some features that affect whether or not a speaker chooses to passivize a verb phrase. We now examine how the average value of each feature changed over time. Note that these are not the estimated coefficients of these features in a model but the observed frequencies of features in the data without considering passivization.

For each decade of the Hansard corpus from 1830, we calculate the average value of each explanatory variable; except for the length features whose values are not Boolean, this average is between zero and one. Figure 2, for example, plots the average value in each decade of the feature indicating whether the agent is local.

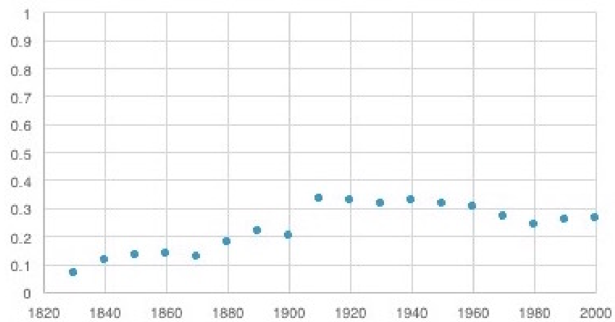

Figure 2: The frequency of local agents has increased over time.

The observed average value of the local agent feature increases over time as shown in Figure 2, and this increase is statistically significant ($p = 0.0003$) if we apply an F-test to the slope of the line of best fit. We apply this test to all the explanatory variables described in Section 3.1. As shown in Table 9, the only ones whose slopes are not significantly different from zero are the agent length feature, the pronoun agent feature, and the indicator for whether the agent is given or new information.

However, not every feature whose slope is significantly different from zero underwent a change as big as the one undergone by the agent feature depicted in Figure 2. In the 1830s, this feature had an average value of 0.069, which means that 6.9% of two-argument verb phrases had a local agent; in contrast, in the 2000s, this same feature had an average value of 0.262, indicating that over 26% of data points had a local agent. The magnitude of this change is thus more than 19 percentage points. For each feature, the raw change in average value from the 1830s to the 2000s is listed in Table 9. (Because agent length and patient length are not binary features, the increase of 0.013 and 0.136 in their average values should not be interpreted on the same scale as the others.)

The indicators for the presence of passives in the preceding verbs and preceding sentences cannot be used to explain the observed decrease in passivization frequency because it is unsurprising that these features decreased in value on average over time as a result of the general declining trend of the rate of passivization.

The relative and wh-clauses are not reliable predictors of passivization according to Table 6 and Table 7; therefore, although their average values exhibit a modest change over time, they are unlikely to be important for explaining the decline of passives.

Feature	Change in Value
Local Agent	0.193
Local Patient	0.014
Pronoun Agent	-0.001 *
Pronoun Patient	-0.042
Agent Length	0.013 *
Patient Length	0.136
Given Agent	-0.004 *
Given Patient	-0.107
Relative Clauses	-0.049
Wh-Clauses	-0.061
Preceding Verbs	-0.182
Preceding Sentences	-0.184

Table 9: Change in average value of each feature between the 1830s and the 2000s (* indicates a statistically insignificant change)

This leaves the locality of the agent, the locality of the patient, the pronoun status of the patient, and the length of the patient as potential explanatory variables for the declining frequency of passivization. (Although the features for local agent and local patient both increased in average value, the change undergone by the latter is a small fraction of that undergone by the former between the 1830s and the 2000s.) To estimate the impact of these explanatory variables, we measure the overall passivization rate when each (binary) feature value is 0 and when it is 1; these rates are listed in Table 10.

Feature	Rate at 0	Rate at 1	Diff.
Local Agent	0.086	0.001	-0.085
Local Patient	0.062	0.125	0.063
Pronoun Patient	0.053	0.153	0.100

Table 10: Passivization rate at different feature values

The passivization rates in the 1830s and 2000s are respectively 8% and 5.2% as illustrated in Figure 1. This means we seek to explain a difference of 2.8% percentage points.

To get an estimate of the contribution to the decline in passivization that can be explained by each feature, we multiply the last column of Table 10 by the changes listed in Table 9. For example, the pronoun status of the patient contributes an estimated 0.42 percentage points to the decline.

We note that the local patient feature actually predicts a slight *increase* in passivization rate, meaning the change is going the wrong way. However, this predicted increase is very small: only about 0.09 percentage points (in comparison, agent locality predicts a 1.64 percentage point decline in the rate).

For the patient length feature, we measure the passivization rate when the value of the feature is 1 and when it is 2 (i.e. when the length is 4, since the feature value is a square root of the actual length). Across all decades, the passivization rate is 0.107 when the value of this feature is 1 and 0.039 when the feature value is 2. This difference of -0.068 should overestimate the effect on passivization rate of increasing the feature value by 1. Multiplying this difference by the change in the average value of this feature gives -0.009, which means that the change in this feature value can explain at most 0.9 percentage points of the declining passivization rate.

If we summed up the effects of the aforementioned contributions, we would seemingly explain the entire 2.8% percentage points. However, because these features correlate with each other, we cannot sum up the estimated effects. In particular, the patient locality, pronoun patient, and patient length features contain overlapping information. However, because one feature alone explains 1.64 percentage points, these features explain well over half of the difference.

5 Changes in Grammar

The significant increase in the frequency of local agents is suggestive, but is the decline in passivization mostly attributable to this lexical choice or, as Mair and Leech (2006) suggested, to increasing convergence with informal speech? We now turn from changes in the average values of features to evidence of changes in grammatical constraints' weights.

5.1 Person Hierarchy

Some languages have a person hierarchy (Aissen, 1999; Bresnan et al., 2001) in which "local" first and second persons outrank "nonlocal" third persons. In one such language, Lummi (Bresnan et al., 2001), the person hierarchy affects passivization in a way such that speakers avoid a construction at all times if the subject is the less prominent argument on the person hierarchy. While no such categorical effects are observed in English, Bresnan et al. use the Switchboard corpus to conclude that statistical preferences for harmonic person-argument associations do exist in English. Our findings are consistent with theirs; moreover, we find that the aforementioned person hierarchy preferences have become stronger in English over time.

The descriptive statistics from two centuries of the British Hansard show the following: With a third-person agent, the passive (instead of the active) is used approximately twice as much with a local patient than with a nonlocal one as shown in Table 11, which is consistent with the explanation that English speakers perceive it to be disharmonic when the subject is less prominent than the object and therefore use the passive to avoid having a nonlocal agent with a local patient in an active construction.

passive percentages	
1/2 acting on 1/2:	0.16%
1/2 acting on 3:	0.1%
3 acting on 3:	7.95%
3 acting on 1/2:	13.94%

Table 11: Descriptive statistics on argument locality

In addition, we find that this gap in frequency has become wider over time: As shown in Figure 3, although this ratio is 1.75 when we consider the entire corpus, it is 1.19 if we focus on the earliest decade.

When we consider a local agent and a third-person patient, the passive is used instead of the active 0.1% of the time across the entire corpus. (This is a weighted average of all the time periods, where a weight is the number of two-argument clauses in a time period.) However, throughout the 1800s, this statistic remained between 0.21% and 0.29%. In contrast, since 1950, this same measure has never been above 0.1%. Moreover, it has dropped from one decade to the next since 1940 without exception. Although the downward trend could be attributed to an overall decline in passive constructions over time, the trend is fairly constant if we instead consider the proportion of passives used with both third-person agent and patient. When we divide the former proportion by the latter, we still observe a clear de-

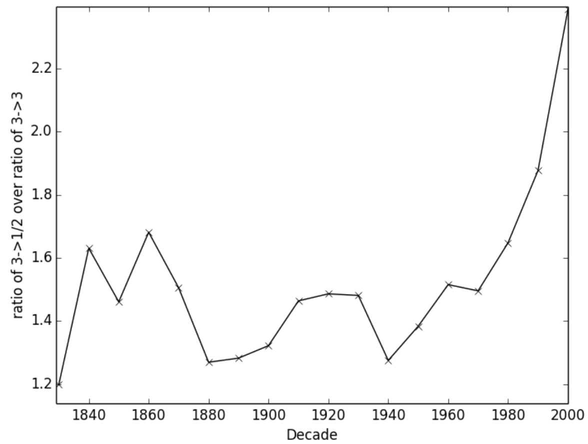

Figure 3: Between years 1830 and 2004: the passive proportion of 3rd-person agents with local patients over the proportion of passives used to express two 3rd-person arguments

cline (as shown in Figure 4), suggesting that English speakers have developed a stronger preference for harmonic person-argument associations over time.

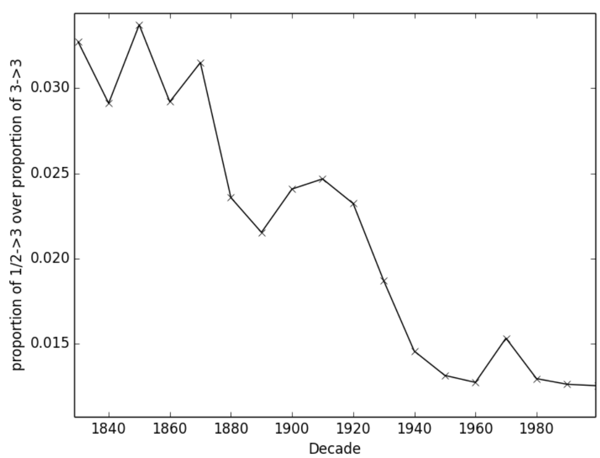

Figure 4: the passive proportion of 1st/2nd-person agent and 3rd person patient over the passive proportion of 3rd-person agent and 3rd-person patient

5.2 Logistic Regression

For each decade starting from 1830, we have created a balanced dataset in which each decade contains 64,000 data points, 50% of which are two-argument passive verb phrases and the other 50% are actives.

We fit a logistic regression model to the data from each decade (independently of the other decades). We learn a different set of coefficients per decade while keeping the features the same across decades.

For the model containing only the person features and no other explanatory variables, the trajectories of the coefficients are consistent with our earlier observations that disharmonic person-argument associations are becoming less preferred over time.

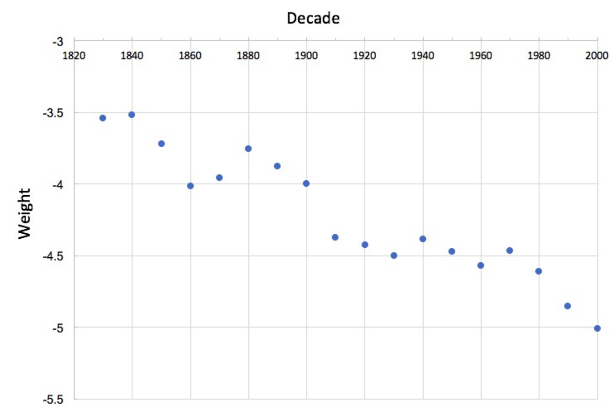

Figure 5: The coefficient of the local agent feature decreases over time.

For example, when the agent is local, we would expect speakers to make it the subject of an utterance by employing the active voice more frequently (and the passive less frequently) over time. Indeed, we observe in Figure 5 that the corresponding person feature's coefficient is decreasing over time.

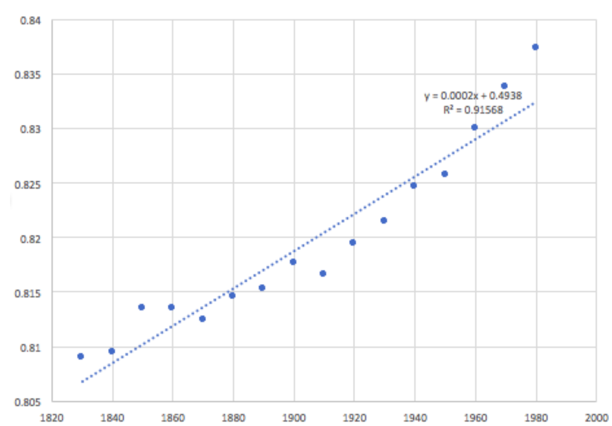

Figure 6: Accuracy of the model with person plus lemma features when trained on various decades and tested on the 1990s (the y-axis was truncated to focus on the change over time)

Figure 6 shows the accuracy achieved by a model with person features and lemma features. The task is predicting passivization for 1990–1999 when trained on data from the decades preceding it. (On the test set, employing the strategy of always guessing the same label achieves an accuracy of 50%.) We see from Figure 6 that performance improves when the training data is from later time periods and that the best performance is obtained by training on the time period immediately preceding the 1990s (i.e. closest to the decade used for testing).

5.3 Hierarchical Model

We have thus far fitted one model per decade assuming different decades are independent. In this section, we instead consider a linear hierarchical model:

$$
\begin{aligned}
a_i &\sim \text{normal}(0, \sigma_a^2) \\
b_i &\sim \text{normal}(0, \sigma_b^2) \\
\theta_i(t) &\sim \text{normal}\left(a_i t + b_i, \sigma^2\right) \\
Y_n &\sim \text{logistic}(X_n \cdot \theta(t))
\end{aligned}
$$

The coefficient vector for the period t is denoted by $\theta(t)$. The labels are denoted by Y_n. Each feature vector is denoted by X_n. Lastly, a_i and b_i are respectively the slope and intercept for the line describing the trajectory over time displayed by feature i. We set σ_a to be 0.1, σ_b to be 5, and σ to be 0.5.

Considering only the two person features, we obtain the following values for a and b using the Stan modeling language (Carpenter et al., 2017; Stan Development Team, 2017) on a dataset of 100,000 two-argument verb phrases consisting of an equal number of data points (sampled at random) per class and per time period (with 2,000 iterations and 4 chains):

$$
\begin{aligned}
a_{localA} &= -0.05, \quad & a_{localP} &= 0.003, \\
b_{localA} &= -6.22, \quad & b_{localP} &= -1.9
\end{aligned}
$$

This model corroborates the findings in earlier sections of the changing constraints against passivizing local agents: In Figure 7, we see that local agent displays a decreasing trend over time, which is consistent with the earlier Figure 5.

6 Conclusions and Future Directions

In this paper, we examined possible causes of the decline in the rate of passivization in English over time. We found that some explanatory variables predict passivization and are themselves changing in value over time; they can therefore be used to explain part of the decline. In particular, the local agent feature increased in average value, and this alone can explain at least half of the decline in passivization. We also found that the person-hierarchy effects noted by Bresnan et al. (2001) became more important over time; the constraint against passivizing clauses containing a local agent gained strength. Moreover, our ablation experiments on the Wall Street Journal and Hansard corpora showed support for the effect on passivization of the structural parallelism observed by Weiner and Labov (1983) and also for the unmarked information structure noted by Halliday (1967) of given information before new information.

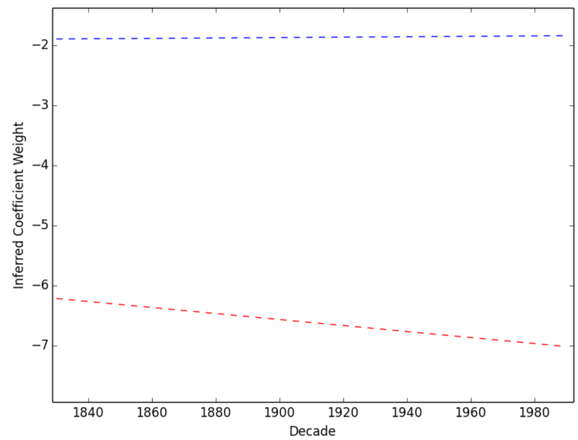

Figure 7: Parameters inferred by the linear model: red corresponds to agents, blue to patients

Future directions include examining animacy as an explanatory variable, which affects passivization in other languages (De Cuypere et al., 2014; Sasaki and Yamazaki, 2006); however, animacy poses difficulties for automated methods as noted by Roland et al. (2007). Prescriptivism is another potential explanatory factor for the changing rate of passivization (Anderwald, 2012). While the prescriptive literature does not directly address the person-hierarchy features and their effect on the passivization rate change, an examination of "awkward" passives in that literature may yield a collection of examples in which first-person agents form the majority. Finally, cohort effects and other speaker variables that might be gleaned from corpora such as the Hansard offer opportunities for sociolinguistic modeling of syntactic change.

References

Judith Aissen. 1999. Agent focus and inverse in Tzotzil. *Language*, 75:451–485.

Marc Alexander and Mark Davies. 2015. Hansard corpus 1803-2005. `http://www.hansard-corpus.org`. Accessed: 2016-03-12.

Lieselotte Anderwald. 2012. Clumsy, awkward or having a peculiar propriety? Prescriptive judgements and language change in the 19th century. *Language Sciences*, 34(1):28–53.

Alexandre Bouchard-Côté, David Hall, Thomas L Griffiths, and Dan Klein. 2013. Automated reconstruction of ancient languages using probabilistic models of sound change. *Proceedings of the National Academy of Sciences*, 110(11):4224–4229.

Joan Bresnan, Shipra Dingare, and Christopher D. Manning. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In *Proceedings of LFG 01*, pages 13–32.

Bob Carpenter, Daniel Lee, Marcus A Brubaker, Allen Riddell, Andrew Gelman, Ben Goodrich, Jiqiang Guo, Matt Hoffman, Michael Betancourt, and Peter Li. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).

Ludovic De Cuypere, Kristof Baten, and Gudrun Rawoens. 2014. A corpus-based analysis of the Swedish passive alternation. *Nordic Journal of Linguistics*, 37(2):199–223.

Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454. Genoa Italy.

T Mark Ellison and Simon Kirby. 2006. Measuring language divergence by intra-lexical comparison. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 273–280. Association for Computational Linguistics.

Michael AK Halliday. 1967. Notes on transitivity and theme in English: Part 2. *Journal of linguistics*, 3(2):199–244.

Geraldine Legendre, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic grammar – a formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pages 388–395.

Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tang, and Martin A Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature*, 449(7163):713–716.

Christian Mair and Geoffrey Leech. 2006. Current changes in English syntax. *The handbook of English linguistics*, 36.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics*, 19(2):313–330.

Mitchell Marcus, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz, and Britta Schasberger. 1994. The Penn treebank: annotating predicate argument structure. In *Proceedings of the workshop on Human Language Technology*, pages 114–119. Association for Computational Linguistics.

Douglas Roland, Frederic Dick, and Jeffrey L Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of memory and language*, 57(3):348–379.

Kan Sasaki and Akie Yamazaki. 2006. Two types of detransitive constructions in the Hokkaido dialect of Japanese. *Passivization and typology: Form and function*, 68.

Stan Development Team. 2017. Pystan: the python interface to stan, version 2.16.0.0. `http://mc-stan.org`.

E Judith Weiner and William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics*, 19(1):29–58.

Christoph Wolk, Joan Bresnan, Anette Rosenbach, and Benedikt Szmrecsnyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica*, 30(3):382–419.